Python Group Project

X3 Group 2

Yilin Yang Yangling Li Tianzi Liu Sichen Du

October 13, 2023

Contents

1. Report Outline 1
1.1 Consumption 1
1.2 Purchasing 1
1.3 Joint Movements 1

2. Data Preperation 1
2.1 Data Cleaning 1
2.2 Data Processing 2

3. Consumption Patterns 4
3.1 Total Consumption Level 4
3.2 Consumption Proportions 5
3.3 Daily Average 5
3.4 Daily Consumption Level 6

4. Purchasing Patterns 8
4.1 Gross Purchase Amount in different gas stations 8
4.2 Gross Purchase Cost in different gas stations 11

5. Relationships between Consumption and Purchasing 14
5.1 Scatter plots and correlation coefficients 14
5.2 Connect joint movements with Discounts 17

6. Conclusions and Recommendations 18

Appendices 20
Appendix A Fuel Cost Per Month at Each Gas Station, 2017-2019 20
Appendix B Geographical distribution of gas stations 24

1. Report Outline

1.1 Consumption

For consumption information, we analyze standing from two different levels: total consumption level and daily consumption level. We create a bar plot indicating total gas and diesel consumption in different gas stations. We use proportion to show relative relationship for gas and diesel. Then break down the data into a daily basis, and plot daily average consumption in different gas stations. Finally, a series of more detailed charts are illustrated, indicating the fluctuation for fuel consumption each day.

1.2 Purchasing

For purchasing information, we analyze the amount and cost of diesel and gasoline purchases at eight service stations respectively in 2017-2019. We present the purchasing trend of eight gas stations on a monthly basis. In the amount of purchases, we summarize from two angles: type of majority of fuel purchase and purchase peak period. Then a series of boxplots are provided to show the statistical characteristics for the cost of purchase each time in different gas stations.

1.3 Joint Movements

Lastly, we aim to analyze the potential relationships between consumption and purchasing behaviors at eight gas stations. The joint movements were visualized by scatter plots, accounting for monthly total fuel consumption and gross purchase cost, with fuel type as the categorical variable. We also computed the correlation coefficient to quantify the strength of relationships. As the final goal for joint movement analysis is to generate meaningful insights for recommendations to gas stations, we included findings of purchase discounts for a holistic view.

2. Data Preparation

2.1 Data Cleaning

The initial step for getting in hand with all these dataframes is to do data cleaning. We utilized 'isna( ).any( )' to check missing values in *df_tank, df_loc, df_fuel* (which is concatenated from *df_fuel1* and *df_fuel2*) and *df_invoice.* In *df_tank* and *df_loc,* there is no missing value. However, there exists missing values in *df_fuel* and *df_invoice.* Hence, we focus on these two dataframes.

(1) Data cleaning for *df_fuel:* This dataframe only contains 2 missing values in column 'Fuel_Level', and we choose to drop them. The reason to drop missing values is that in our later analysis, we only used the beginning and ending fuel level for calculating daily consumption. Hence, the 2 missing values in Fuel_Level can be dropped without influencing our analysis results. (If there are

missing values for the beginning and ending data for different days, then we may need to use forward-fill or backward-fill)

(2) Data cleaning for *df_invoice:* This dataframe contains 41 missing values in column 'Invoice ID', 42 missing values in column 'Gross Purchase Cost', 'Amount Purchased' and 'Fuel Type'. We choose to drop these missing values for two reasons. First, invoice data refers to the purchasing detail. For missing values, there is no reason for us to use forward-fill or backward-fill to 'create' a new invoice record; Second, location of rows for missing values are the same. This means for records with any one missing value, the rest of other variables are also missing. Hence, the record for us is meaningless. We also don't fill them with 0 since in later analysis we will use left join for consumption information and invoice information, and we will fill missing values for invoice with 0 at that time.

(3) We use df.merge to organize the data we need into a list. Sort and rename the list to create a more visual arrangement. For example, use the left_on and right_on directives to extract the data you need from the original csv file. Reorder the table by the number of the gas station and the chronological order of the bill. For the following chart presentation.

2.2 Data Processing

After cleaning, we start to process the data and calculate daily fuel consumption in gas and diesel type. Our logic can be divided into two steps: (1) calculate the difference from beginning to ending on each day, the difference is daily consumption on that day; (2) then we take invoice data into consideration. For those days when there are invoice records, we add invoice back to get true daily consumption.

This can be illustrated with the table below:

Table 2.1 Illustration of calculating true daily consumption

Fuel Level (Beginning)	Fuel Level (Ending)	Daily consumption (no invoice)	Invoice on that day	True daily consumption
200	100	100	300	400

As in *Table 2.1*, if there is no invoice record on that day, then daily consumption can be calculated simply by using "Beginning Fuel Level - Ending Fuel Level = 100". However, since invoice on that day is 300, then true daily consumption is "Beginning Fuel Level + Invoice – Ending Fuel Level", which equals to 400.

With this logic, we create a lambda function to calculate the difference between beginning and ending Fuel Level for each Tank_ID on each day. Since invoice data is on a gas station level, we calculate the total daily consumption for all tanks (gas or diesel) on a gas station level.

For invoice details, we group by date and location to get the result for total amount of purchasing in each fuel type, each gas station and each day. Since invoice records are less than consumption records, we use left join to add invoice detail to daily consumption data. Obviously, there would exist lots of missing value for invoice information because not every gas station will purchase fuel every day. For these NaN values, it means there are no invoice on that day, this equals to invoice on that day equals 0. Hence, we use fillna(0) to deal with these missing values.

Then we calculate the true daily consumption by adding back to the invoice. However, there still exists some negative value for total consumption. For example, the beginning fuel level is 100, the ending fuel level is 400, but there is no invoice on that day. For these anomalies, our explanation is that the criteria for separating each business day is different. In our calculation, we define that the very beginning of each day is at 00:00:00, and ends at 23:59:59. However, gas stations might define the opening time as the beginning and closing time as the ending on each day. This can cause differences in defining the records for a specific day, which in turn affect calculating true daily consumptions. We don't know the exact criteria for selecting days for these gas stations, so we still use the criteria that "starts on 00:00:00, ends on 23:59:59" for analysis.

After tagging each day with a date_number and adding gas station name to the corresponding gas station location ID, we finally obtain *df_total_cons* for analyzing consumption patterns, as shown in *Figure 2.1* below.

	date_number	Location	Gas Consumption	Gas Amount Purchased	Total Gas Consumption	Diesel Consumption	Diesel Amount Purchased	Total Diesel Consumption	Timestamp	Gas Station Name
0	0	1	11637.0	0.0	11637.0	5898.0	0.0	5898.0	2017-01-01	EastMount
1	0	2	1669.0	0.0	1669.0	1821.0	0.0	1821.0	2017-01-01	Eastgate
2	0	3	0.0	0.0	0.0	4.0	0.0	4.0	2017-01-01	Central
3	0	4	-15.0	0.0	0.0	-3.0	0.0	0.0	2017-01-01	Chedoke
4	0	5	8.0	0.0	8.0	8.0	0.0	8.0	2017-01-01	Mountain View

Figure 2.1. A snapshot of df_total_cons

3. Consumption Patterns

After obtaining the DataFrame *df_total_cons*, we are able to perform data visualization for further analysis as follows: total consumption level at each gas station, the daily average amount of fuel consumed, and the actual number in liters consumed per day.

3.1 Total Consumption Level

Firstly, we plan to explore the consumption pattern for different gas stations in a more general way. By getting the total consumption for diesel and gas can provide us with a clear picture of the scale and relative consumption amount for different types of fuels. Hence, we summarize all records in *df_total_cons* for each gas station on different days and calculate the sum for 'Total Gas Consumption', 'Total Diesel Consumption'. Based on this, we created a stacked bar plot (*Figure 3.1*) showing the exact scale and value for total consumption for different fuels.

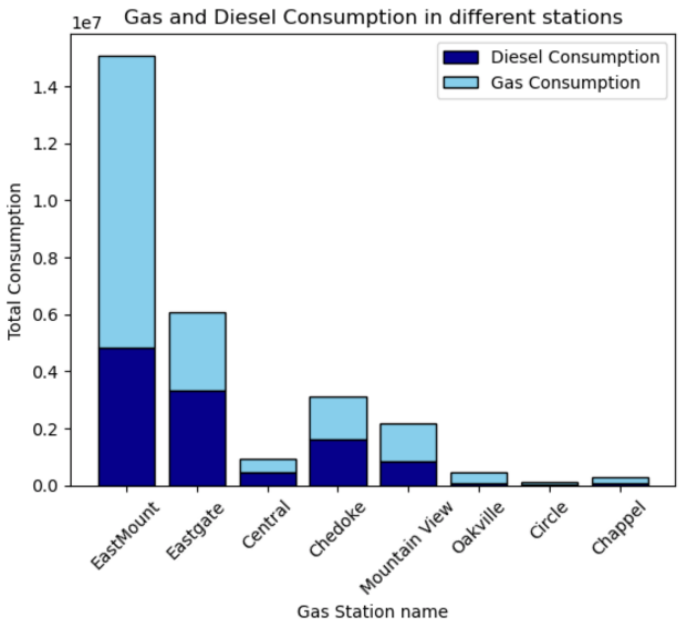


Figure 3.1 Gas and Diesel Consumption in different stations

According to *Figure 3.1*, the gas station name is listed as x-label, while the y-axis shows the value for total consumption. The bars of 'Gas Consumption' are stacked over Diesel Consumption. As we can see, gas station 'EastMount' has the highest consumption volume for both diesel and gas, while gas station 'Circle' has the least consumption volume. Another point worth noticing is the relative amount of consumption for different types of fuels. For gas stations 'EastMount', 'Mountain View', 'Oakville' and 'Chappel', the consumption for gas is significantly larger than that for diesel. While for gas stations 'Eastgate' and 'Chedoke' we see that diesel consumption is slightly bigger than that for gas. In summary, the total volume for 'EastMount' is no doubt the highest among all gas stations. We can conclude that this gas station is the busiest and might need to be refilled frequently.

3.2 Consumption Proportions

Next is the consumption detail for different types of fuel. In the original dataset, we have 'U', 'P' and 'D' types of fuel, while 'U' and 'P' can both be classified into 'Gas' type.
To investigate the exact proportion for gas and diesel among gas stations, we calculate 'Total Consumption' as the addition of 'Total Gas Consumption' and 'Total Diesel Consumption', and regard this amount as 100%. Then we express 'Gas proportion', 'Diesel proportion' as a percentage of total consumption in *Figure 3.2*.

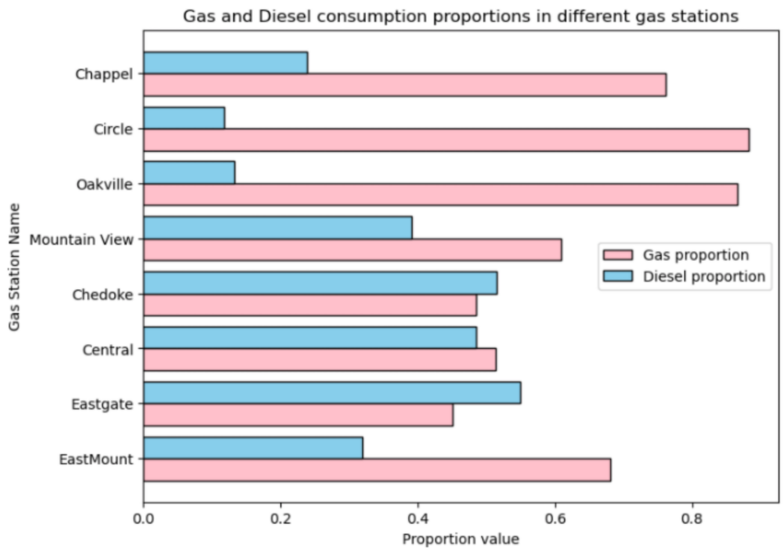


Figure 3.2 Gas and Diesel consumption proportions in different gas stations

Figure 3.2 exhibits a clear and straightforward comparison between gas and diesel. Among 8 gas stations, 6 of them have higher gas proportion, and 2 of them have higher diesel proportion. Besides, in general gas proportion has a higher level, with a maximum value nearly 90%. However, the maximum value for diesel proportion is less than 60%.

3.3 Daily Average

After reviewing the total consumption of different fuel types, we further explored the daily averages at each gas station. First, we constructed a new data frame grouping by gas station locations and calculated the average consumption by total consumption divided by total number of days. Next, we utilized a side-by-side barplot to interpret the data frame. As shown below, the x-axis recorded the names of eight gas stations in total while the y-axis traced the average daily consumption for each fuel type respectively.

In *Figure 3.3*, East Mount gas station ranked first (with 11,245 liters of gas and 5,263 liters of diesel consumed per day) among all eight gas stations when comparing daily amount of fuel consumed for both gas and diesel, followed by Eastgate and Chedoke. A reasonable explanation could be these three gas stations operated at relatively larger scales and thus became more popular among the locals. The number of tanks and the corresponding capacity from *df_tank* data frame also validates this point, as East Mount had 6 tanks in total and Eastgate owned 4 tanks (with one tank, T16, that had nearly doubled capacity of 70,000 liters compared to other tanks capacity around 30,000 liters to 40,000 liters). Similarly, Circle gas station only possessed 2 tanks with 5,000 liters capacity each. Therefore, Circle gas station was built as a small business and the supply of fuel was limited, confirming the low average daily consumptions as well.

In addition, there is no clear pattern of fuel type preferences for gas stations in general. As some gas stations like East Mount and Oakville consumed more gas than diesel on average for a single day, there exists other consumption patterns. For example, the Central gas station has about the same level of daily consumption amount for gas and diesel. And Eastgate on the other hand had more diesel consumed daily than gas on average.

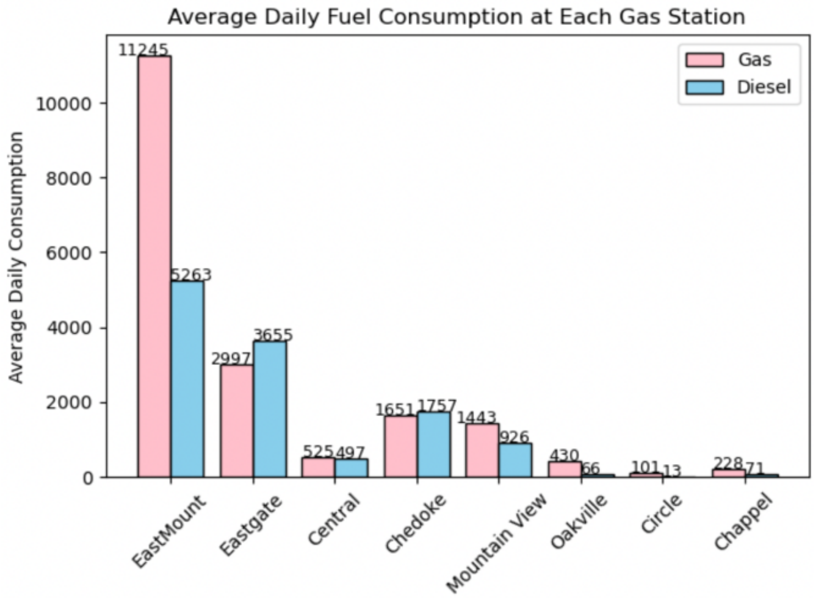


Figure 3.3 Average Daily Fuel Consumption at Each Gas Station

3.4 Daily Consumption Level

Moreover, we also examined the actual amount of fuel consumed every day. The line charts were formed for each gas station separately to track gas and diesel consumption across time intervals (*Figure 3.4*). As previously mentioned, the daily consumption was calculated based on the assumption of natural divisions of time, which means we grouped all the records of fuel level between 00:00:00 a.m. till 23:59:59 p.m. into one day, regardless of each gas stations' working hours. The tradeoff of this method was some rarely happened negatives in the total consumptions, which were replaced by 0 in this case for plotting line charts.

The x-axis accounted for time intervals, ranging from 2017/01/01 to 2019/08/15. Since there were 914 days in total, we decided to display 4-month intervals on x-axis labels, thus placing 2019/09 at the end of labels. Besides, the two y-axes were recognizing fuel consumption level, with the left axis for gas and the right axis for diesel. We adopted two y-axes because gas stations experienced gas consumption and diesel consumption at various magnitudes. And different scales helped us to observe how the line fluctuated overtime.

Analyzing the movement of lines, it is clear that sharp increases occurred occasionally, indicating an abrupt surge in demand on a single day. These peaks have appeared in all gas station consumption patterns and thus we believe adequate storage of additional fuel supply is critical. The ability to predict for such changes will impact the profitability of gas stations, as well as adjustments to regular replenishment plans. Yet, more data is needed for modeling future consumption patterns. Potential independent factors like weather of the day or time of vacations should be considered. The lack of advanced skills of partitioning data and building appropriate linear regression models via Python is another limit we faced.

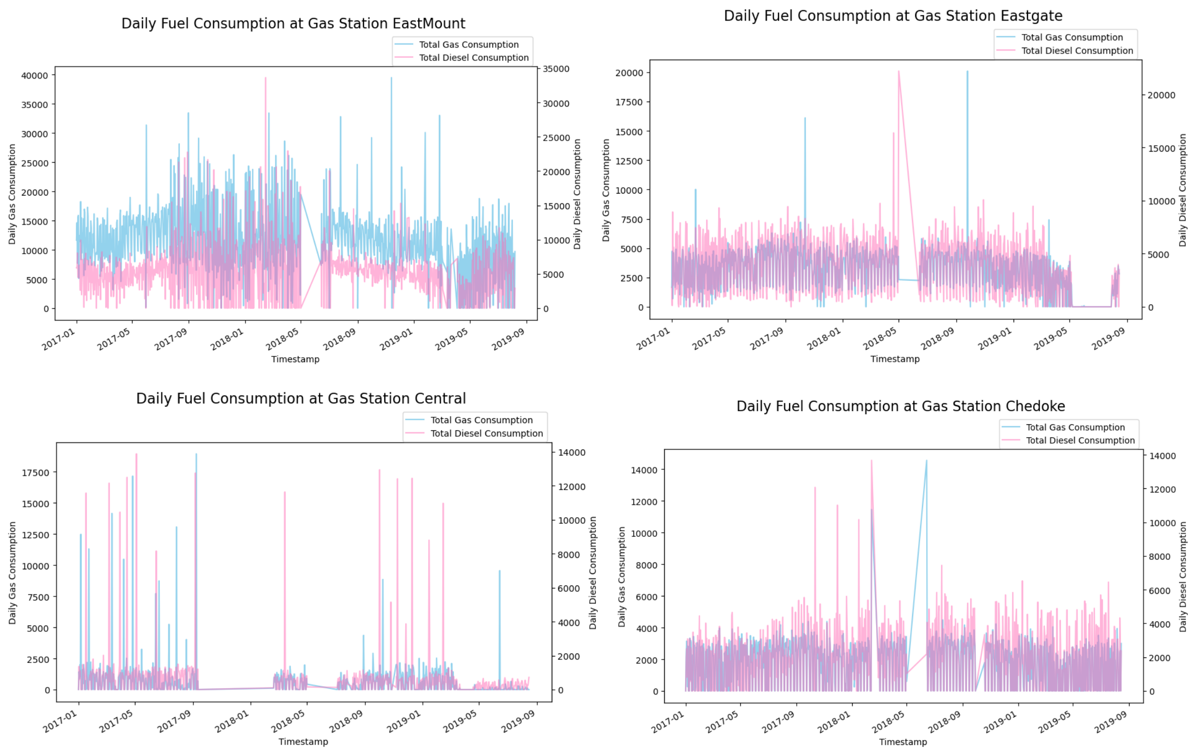


Figure 3.4 Daily Fuel Consumption Patterns at Each Gas Station

Figure 3.4 (Continued) Daily Fuel Consumption at Each Gas Station

4. Purchasing Patterns

4.1 Gross Purchase Amount at different gas stations

(1) **More gas purchases, peak from June to August.** We draw the purchasing trend of gas and diesel in EastMount in each month from 2017 to 2019. As shown in *Figure 4.1*, gas purchases are higher than diesel throughout the year. Besides, for both fuel types, the largest number of purchases are made from June to August, while the smallest number of purchases are from October to December.

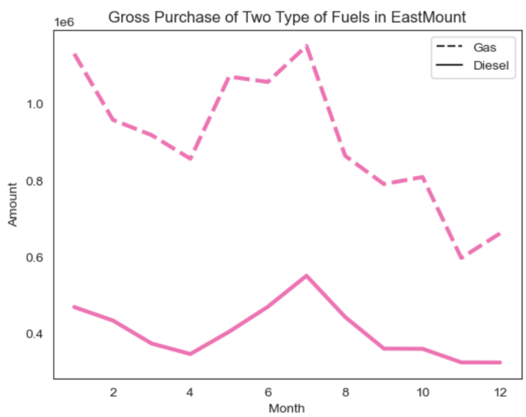


Figure 4.1 Gross purchase for different types of fuels in EastMount

The same results can be obtained from *Figure 4.2* and *Figure 4.3*. The two graphs show the amount of gas and diesel purchases from 2017 to 2019. General trend continues to show the highest purchases of gas and diesel are made during the June and August of the year. However, two figures, *Figure 4.2* and *Figure 4.3*, are the most representative of all gas stations. They give the clearest feedback about changes in the data over months and are closer to the previous line chart.

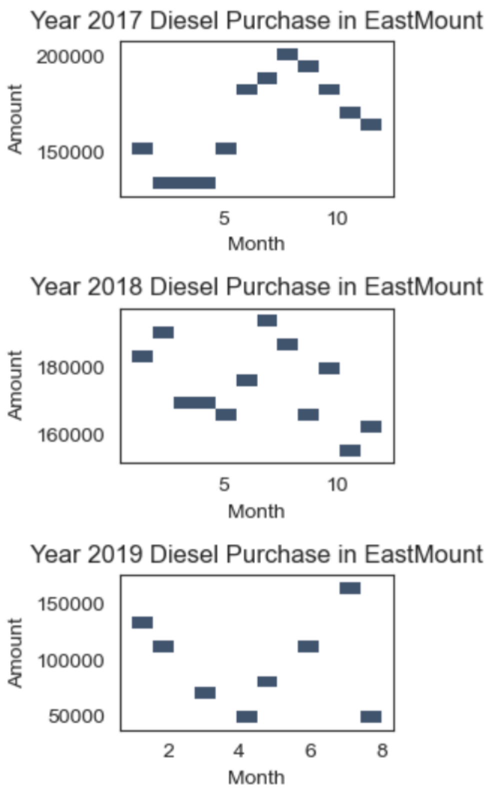


Figure 4.2
Monthly Diesel Purchase for EastMount

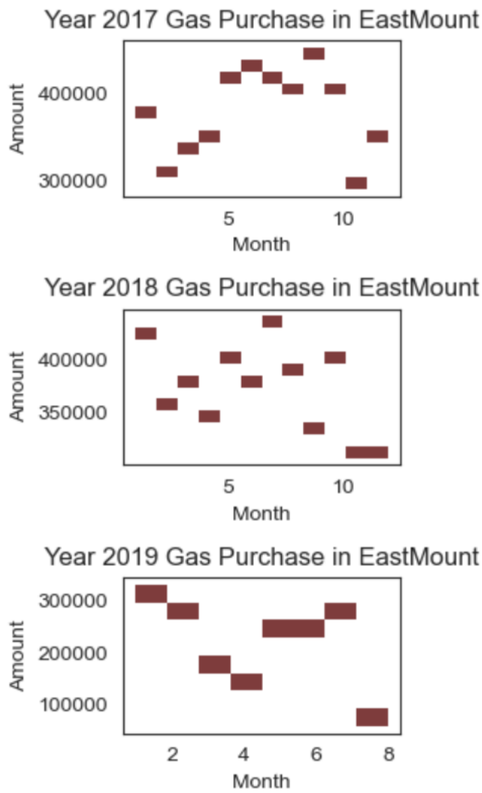


Figure 4.3
Monthly Gas Purchase for EastMount

(2) **More diesel purchases, peak from January to February.** From *Figure 4.4,* we can draw the purchasing trend of diesel and gas in EastGate and Mountain in each month of the year from 2017 to 2019. Eastgate purchases more diesel than gas throughout the year. And Mountain purchases more gas than diesel throughout the year. However, for diesel and gas, the number of purchases was the highest in January-February and then showed a downward trend.

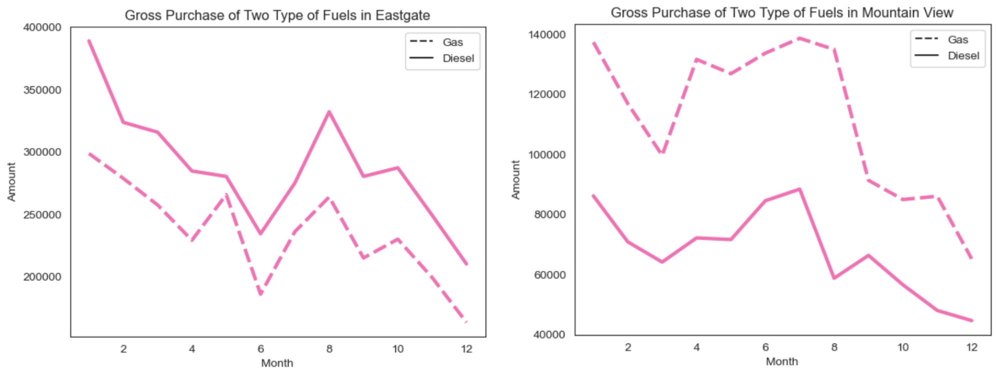


Figure 4.4 Two Type of fuels Gross Purchase for Eastgate and Mountain

(3) **No specific preferred type, peak period varies from time to time.** From *Figure 4.5*, we can draw the purchasing trend of diesel and gas in Central and Chedoke in each month of the year from 2017 to 2019. Their monthly purchases and the comparison do not show a clear pattern. They purchase similar amounts of diesel and gas each month. Sometimes gas purchases outweigh diesel, while sometimes on the contrary.

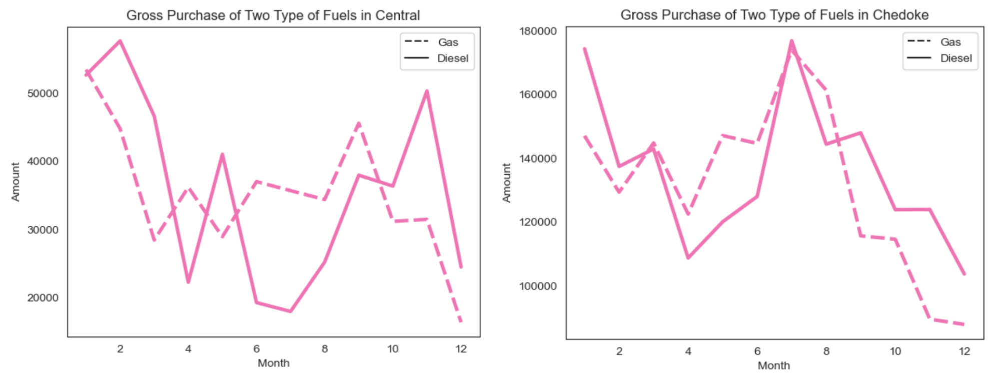


Figure 4.5 Two Type of fuels Gross Purchase for Central and Chedoke

(4) **Diesel data is missing for some months.** *Figure 4.6* shows the purchasing trend in Oakville, Circle and Chappel in each month of the year from 2017 to 2019. Overall, their purchasing patterns for Oakville and Circle are similar, with more gas purchases throughout the year. However, while station Chappel sold more gasoline in the first half of the year, the purchases of diesel increased significantly from May and are close to the purchase volume of gas in the second half of the year.
Besides, the three stations all have diesel records missing. Circle has no diesel record for January to March; Chappel has no diesel record in January to February; While Oakville has no diesel record in October to December.

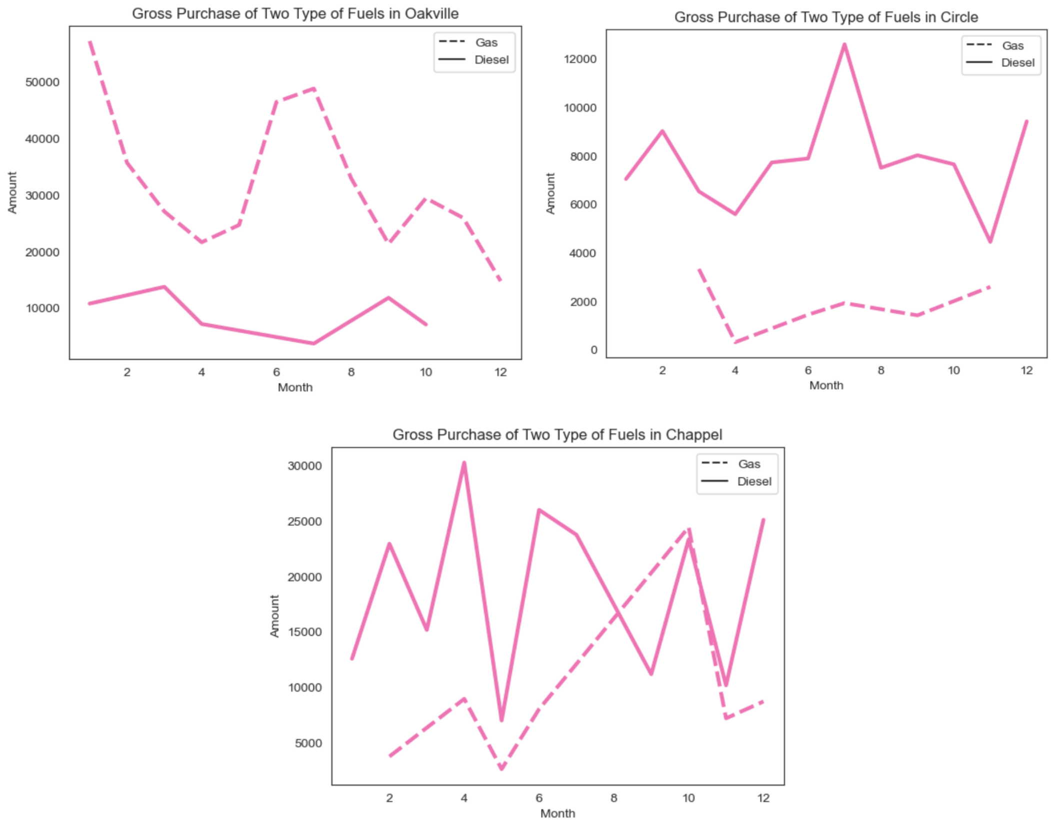


Figure 4.6 Two Type of fuels Gross Purchase for Oakville, Circle and Chappel

4.2 Gross Purchase Cost in different gas stations

The following boxplots (*Figure 4.7*) visualize the distribution of spending per purchase of gas or diesel for each Station over the three-year period 2017-2019.

(1) *Figure 4.7(a)* shows that EastMount's median spend when acquiring gas in a single purchase over the three-year period 2017-2019 is about $11,000. Similarly, the median amount spent on a single purchase of diesel is around $10,000. However, comparing the costs of purchasing gas and diesel shows that EastMount's costs for gas are more spread out and wider, with 50% of the data spread between about $8,000 and $28,000.

On the other hand, the costs for diesel are more concentrated, with 50% of the costs falling in the range of $9,000 to $11,000. At the same time, however, the box plot depicting diesel has a number of outliers, which suggests that EastMount occasionally purchases far more diesel than it has in the past, which may be related to its higher-than-usual consumption of diesel at particular times. An in-depth analysis of the cause of this outlier would need to be analyzed in the context of the reality of the market, where we could also consider the inflation factor.

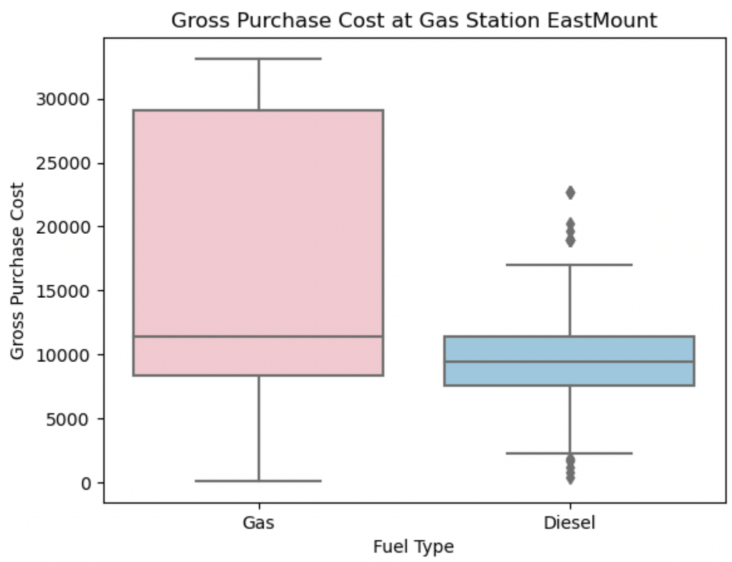


Figure 4.7 (a)

(2) From *Figure 4.7(b)* we can see*,* similar to EastMount, Central has a similar pattern when purchasing gas and diesel. Compared to the diesel purchase range, the cost of gas purchases is much broader.

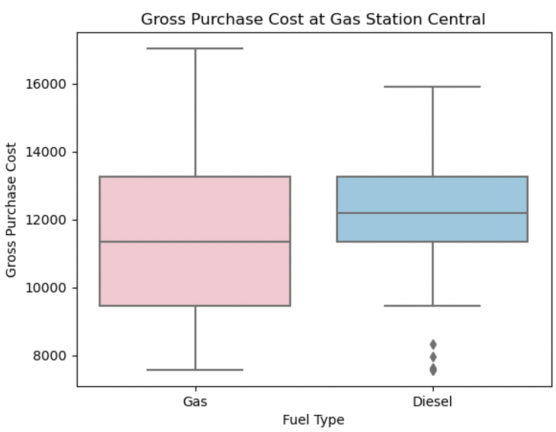


Figure 4.7 (b)

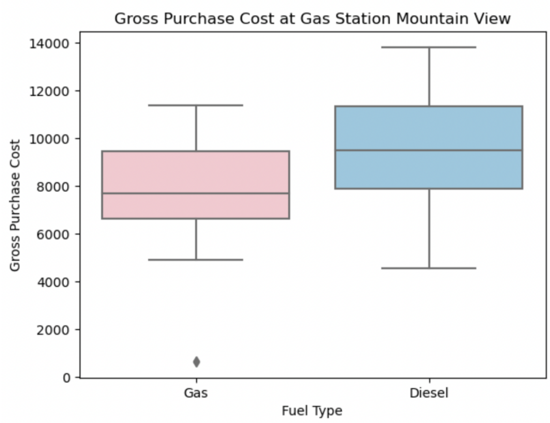


Figure 4.7 (c)

Then, observing the distribution of Mountain View's costs (*Figure 4.7 (c))* in purchasing gas and diesel reveals a pattern with similarities to EastMount and Central: The median values of cost for a single purchase of gas and diesel are relatively similar. The median cost of a single purchase of gas is around $8,000, while the median cost of a single purchase of diesel is around $10,000.

Nevertheless, the difference between EastMount (*Figure 4.7 (a)*) and Station Central (*Figure 4.7 (b)*) is that the distribution of single purchases of gas in Mountain View is more concentrated, and Mountain View's diesel purchases have no outliers. This reflects that Station Mountain View has stable gas and diesel consumption, resulting in a more stable purchase volume and cost.

(3) Observation of *Figure 4.7 (d) - (g)* indicate that the four stations, Eastgate, Chedoke, Chappel and Oakville, all share a similar pattern in terms of the cost of purchasing gas and diesel, exhibiting the following characteristics:

- Whether purchasing gas or diesel, the cost of a single purchase is concentrated in a small interval, above and below $10,000.
- There are a number of outliers that deviate from the usual cost.

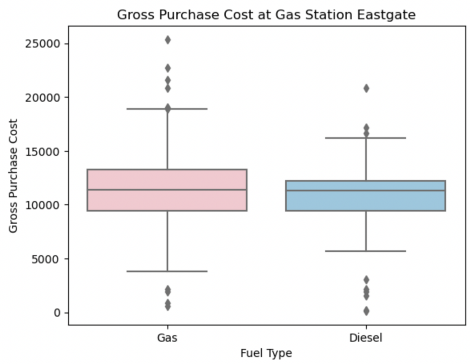


Figure 4.7 (d)

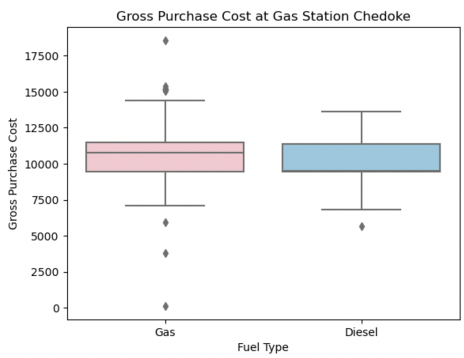


Figure 4.7 (e)

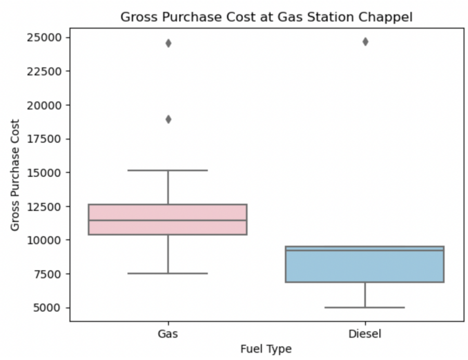


Figure 4.7 (f)

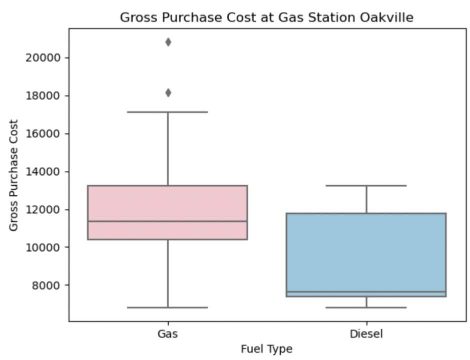


Figure 4.7 (g)

Slightly different from the other two gas stations, Chappel's diesel purchases are almost exclusively centered on a narrow range. Except for one exception of an outlier at $25,000. The cost for its single diesel purchase ranges from a minimum of about $5,000 to a maximum of about $10,000.

In addition, compared to the other two stations, the cost of a single purchase of diesel is more significantly lower than the cost of a single purchase of gas in Chappel and Oakville, reflecting a greater focus on sales of gas in Chappel and Oakville.

(4) Lastly, we take a look at the station Circle. As shown in *Figure 4.7(h),* Circle's costs for single purchases of both gas and diesel are very low, all under $3000. This may be due to two reasons:

- Circle is a very small gas station, with fewer customers and fewer sales, and thus less need for heavy restocking.
- Circle adopts a model of frequent stocking and low single stocking.

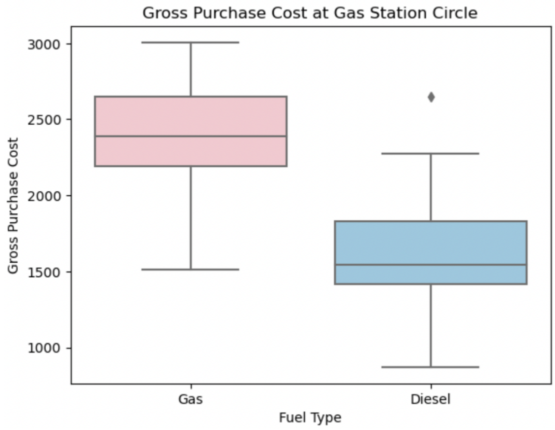


Figure 4.7 (h)

Figure 4.7 Gross Purchase Cost at Each Station

In addition, Circle has a similar distribution of costs when purchasing gas and diesel, with 50% of the costs concentrated in the interval of about $1,000 in length and 100% of the costs concentrated in the interval of $1,500 in length. The difference is that, overall, Circle spends significantly more money on gas than diesel in a single purchase.

5. Relationships between Consumption and Purchasing

5.1 Scatter plots and correlation coefficients

Following the individual analysis of consumption patterns and purchasing patterns, we further explored the possible relationships between consumption and purchasing behaviors at a single gas station. In order to link the data between consumption and purchasing, several steps of adjustments have been applied to form a new data frame called *df_monthly_joint*:

(1) Use timestamps as the foundation to mark each row with a distinct month and year value.
(2) Calculate the monthly fuel consumption level for gas and diesel respectively at each gas station.
(3) Calculate the monthly gross cost and amount of fuel purchased at each gas station.
(4) Merge the above data frames into *df_monthly_joint* according to month, year and gas station locations.

Based on *df_monthly_joint*, a scatter plot was constructed by monthly fuel consumption on the x-axis and monthly fuel purchase cost on the y-axis. The blue dots represent the data spread of gas while the pink dots represent the data spread of diesel. And each gas station was plotted independently.

Overall, the scatter plots of different gas stations demonstrated a general positive linear relationship between total consumption and gross purchasing cost. To further quantify the strength of correlation at each gas station, we built a for loop to loop through gas stations, and then calculate the correlation coefficient between consumption and cost. From the correlation coefficient listed in *Table 5.1*, it is clear that most of the correlation is positive, with only one negative at Oakville gas station for diesel type of fuel. Yet, -0.021 is close to 0 and can be considered as no relationships were found between consumption and cost. We believe this is because Oakville seldomly made purchases and the invoice data sample was too small to find the true pattern. Still, the majority of gas stations showed either a strong or moderate level of positive relationships for monthly consumption and cost.

Table 5.1 Correlation Coefficient between Consumption and Cost

	EastMount	Eastgate	Central	Chedoke	Mountain View	Oakville	Circle	Chappel
Gas	0.566	0.767	0.659	0.435	0.579	0.530	0.382	0.020
Diesel	0.537	0.828	0.639	0.549	0.410	-0.021	0.142	0.022

Below is the scatter plot for the Eastgate gas station (*Figure 5.1*). We selected Eastgate as a typical example to focus on since the dots seemed to gather closely to form a trendline.

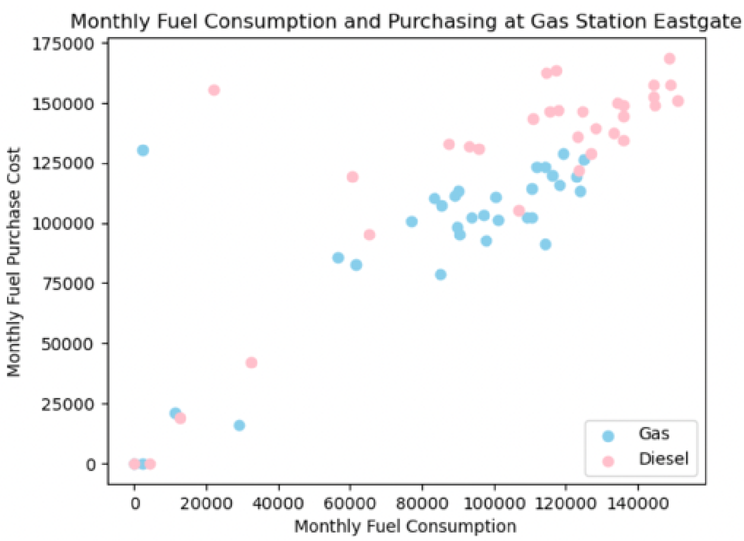


Figure 5.1. Monthly Fuel Consumption and Purchasing at Eastgate Gas Station

As we can see in *Figure 5.1,* correlation coefficient absolute values for gas and diesel were also the largest among all gas stations, suggesting strongest positive relationships. The positive relationship indicated an increase in fuel consumption is related to an increase in fuel purchase cost. Intuitively, it is reasonable to argue that as a gas station consumed more fuel in a certain month, it would also order more fuel to back up supply, leading to the increase of total purchase cost.

Another example we noticed is *Figure 5.2*, the scatter plot for the Chappel gas station. Unlike the Eastgate gas station, the dots seemed to spread more across the graph without a clear trendline that can be perceived. In this case, the relationship between fuel consumption and purchasing cost is weaker compared to the Eastgate gas station. The correlation coefficients were also 0.020 and 0.022 respectively, which means no relationship was found.

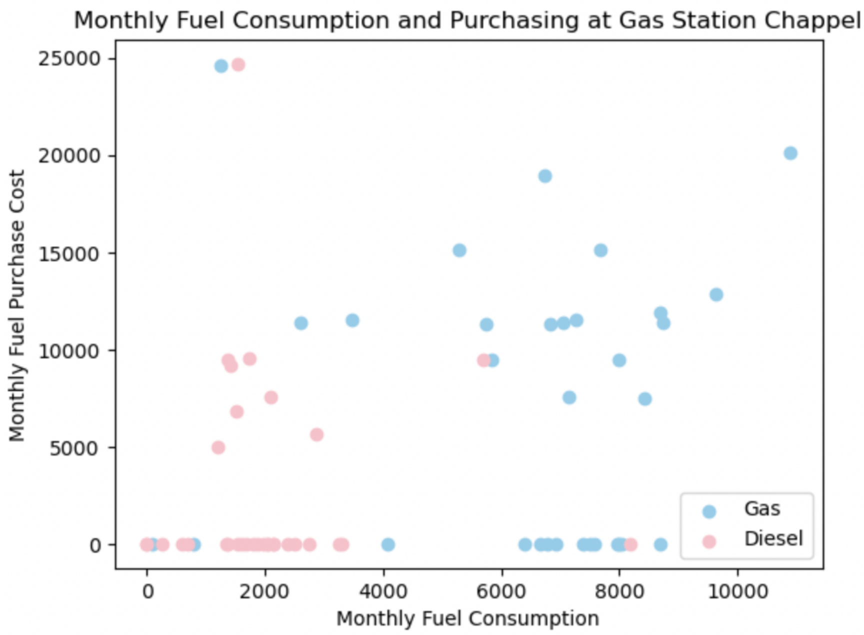


Figure 5.2. Monthly Fuel Consumption and Purchasing at Chappel Gas Station

We interpreted those weaker relationships as postponed purchasing behaviors at specific gas stations. Since we aggregated the consumption level and cost into one-month units, high consumption level with low purchase cost like the dots falling on the horizontal axis means the gas station didn’t respond to the changes in fuel consumption level in a timely manner. Though the demand of fuel was high in a month, the gas station didn’t prepare more fuel as supply in turn to acknowledge the movements in fuel level.

After checking the tank capacity for Chappel gas station, we found Chappel owned two tanks at 40,000 liters of capacity each. Since the amount of consumption for a month hardly exceeded 10,000 liters, Chappel apparently didn't need to purchase more fuel every month, contributing to those dots with 0 purchase costs. Thus, deciding on the effectiveness of the purchasing strategy at Chappel gas station still requires a more advanced and comprehensive approach, taking factors such as discounts into account.

5.2 Connect joint movements with Discounts

In this part, we obtain the discount type when selling diesel and gas at each gas station. *Figures 5.3* are the three most representative tables among all gas stations. *Figure 5.3(a)* is a graph of EastMount gas station. It can be seen that although the 'No discount' category still accounts for the majority of all diesel and gas purchases, 1.17% of diesel and 11.42% of gas purchases fall in the discount type 2, which is 2 cents per liter. At the same time, there are 23.23% gas purchases with discount type 3 (3 cents per liter).

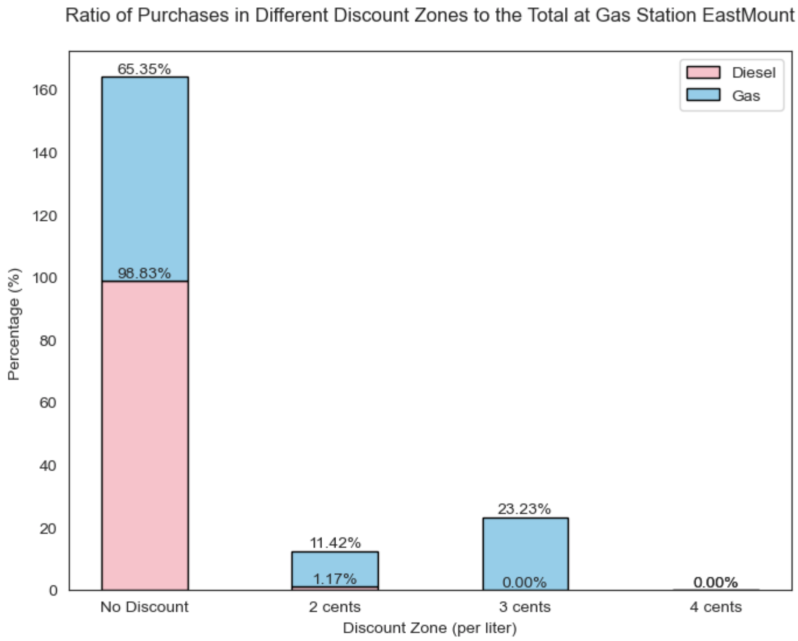


Figure 5.3 (a) Ratio of Purchases Discount in EastMount

Next, shown in *Figure. 5.3 (b)* below is the discount detail in Chappel gas station. It can be seen that although the 'No discount' category accounts for the majority of all diesel and gas purchases, 11.11% of diesel and 10.53% of gas purchases for 2-cents discount. But there is no diesel or gas purchase that meets the purchase quantity of 25,000 liters and above for higher discounts.

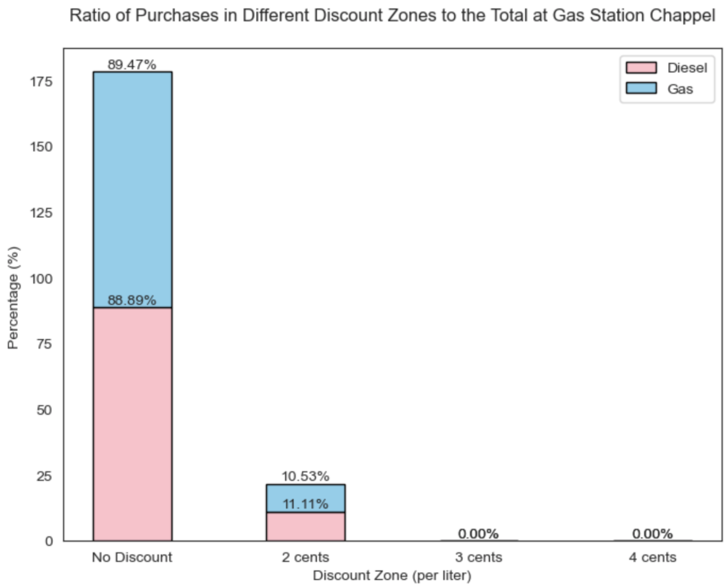


Figure 5.3 (b) Ratio of Purchases Discount in Chappel

Figure 5.3 (c) is a graph of the Mountain View gas station. This is the most extreme case. 100% of diesel and gas purchases are settled without any discounts. There is no diesel or gas purchase in discount types 2, 3 and 4 categories.

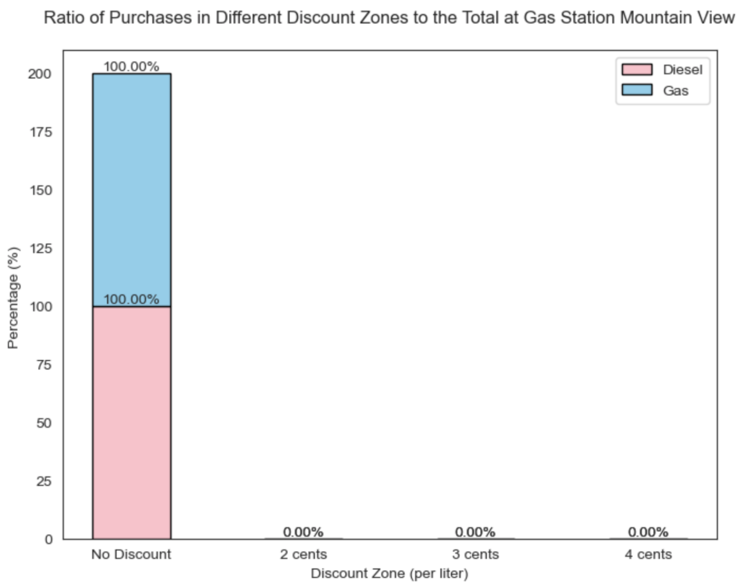


Figure 5.3 (c) Ratio of Purchases Discount in Mountain View

6. Conclusion and Recommendations

From the observation above, the gas stations were not capable of achieving higher discounts per liter because of the low purchase quantity. Even though factors like the size of the gas station (tank capacity) and the consumption amount should be considered, which small gas stations potentially would purchase less compared to larger stations, we suggest all gas stations to reduce purchase frequency and increase the amount purchased for each transaction. Yet, quantitative resolutions can not be formed due to lack of information. For example, as gas stations would prefer larger tanks to store more fuel at each purchase, the cost of tanks at various capacities, or the storage and obsolescent cost for fuel purchased would also affect the ultimate resolution.

Moreover, the limitations in our ability to construct regression models in Python was another aspect. For future studies, we could adopt data partitioning and build predictive models for fuel consumption. And thus the purchasing behavior linked with consumption will be calculated. When the fuel level in tanks becomes close to zero, it means the gas station should purchase the appropriate type of fuel in advance. The replenishment frequency and quantity will be determined by the cost discounts provided accordingly. Other recommendations for tank capacity, operation and maintenance costs also need to be addressed to fully promote the profitability of gas stations.

In conclusion, the consumption pattern varied a lot for each gas station we investigated. Some gas stations consumed more gas while others consumed more diesel. The daily amount consumed also differed in magnitudes, ranging from 11,245 liters to 13 liters on average. On the other hand, the purchasing pattern was more consistent throughout the 3 years length. Most gas stations didn't receive any discounts for most of their purchases, with purchase quantity less than 15,000 liters. And positive linear relationships could be established between monthly gross fuel purchase cost and fuel consumption.

Appendices

Appendix A Fuel Cost Per Month at Each Gas Station, 2017-2019

In our report, the fuel cost per month in station EastMount is shown as *Figure 4.2* and *Figure 4.3* in Section 4. The rest of 7 stations are listed as appendix.

By looking at how much each gas station spends on diesel and gas, respectively, for each month from 2017 to 2019, it can be seen that the same gas station purchases diesel and gas each year in a relatively similar pattern. For example, when Central purchased diesel in 2017-2019, it always spent more in February-April and less around December-February. Other gas stations show similar characteristics.

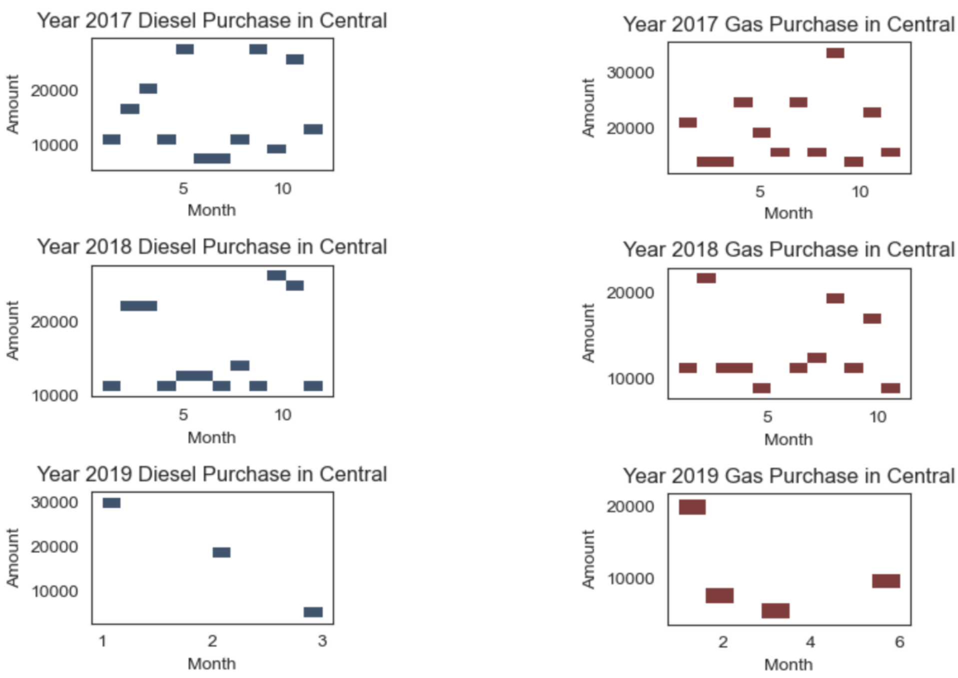


Figure A.1 Diesel and Gas Purchase in Central

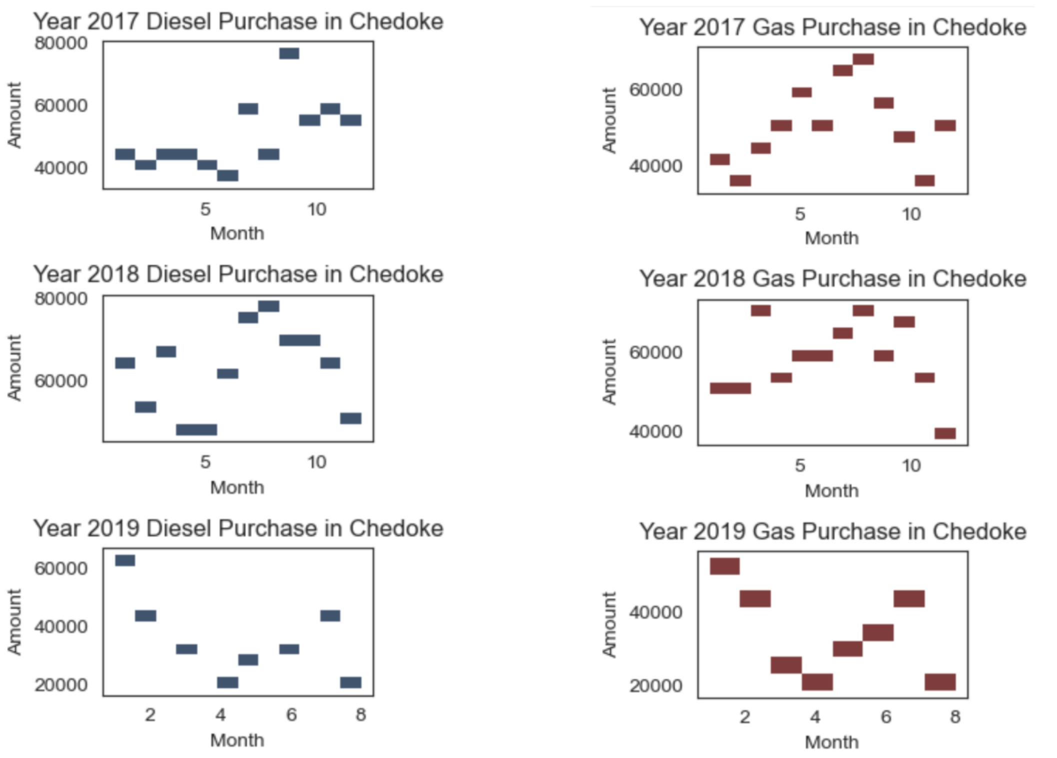


Figure A.2 Diesel and Gas Purchase in Chedoke

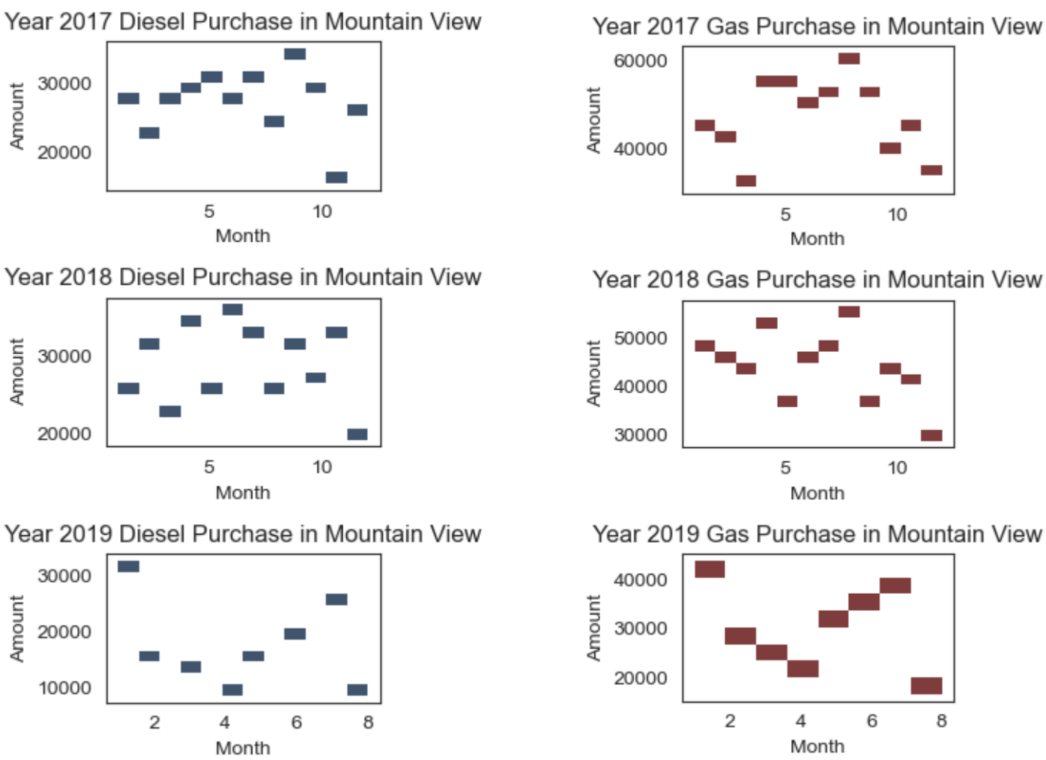


Figure A.3 Diesel and Gas Purchase in Mountain View

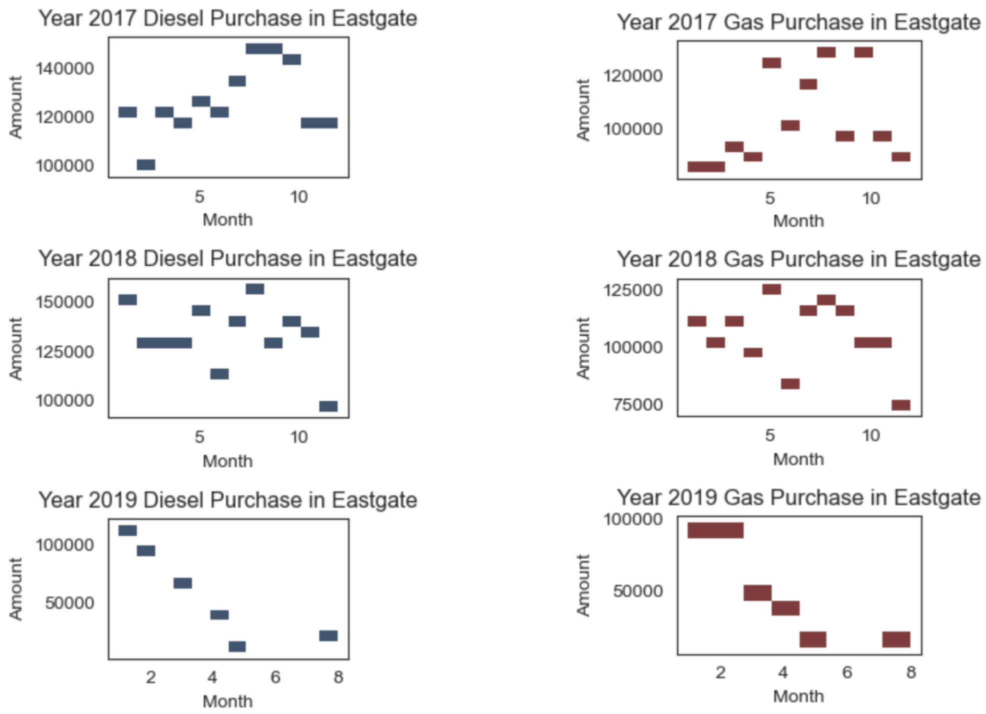


Figure A.4 Diesel and Gas Purchase in Eastgate

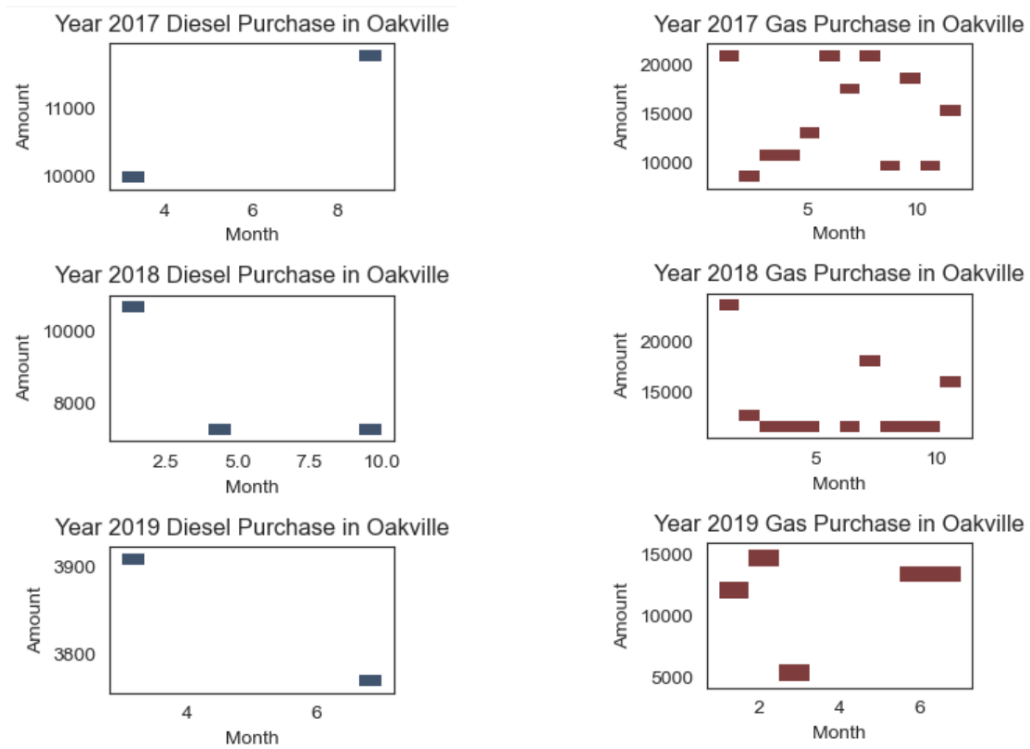


Figure A.5 Diesel and Gas Purchase in Oakville

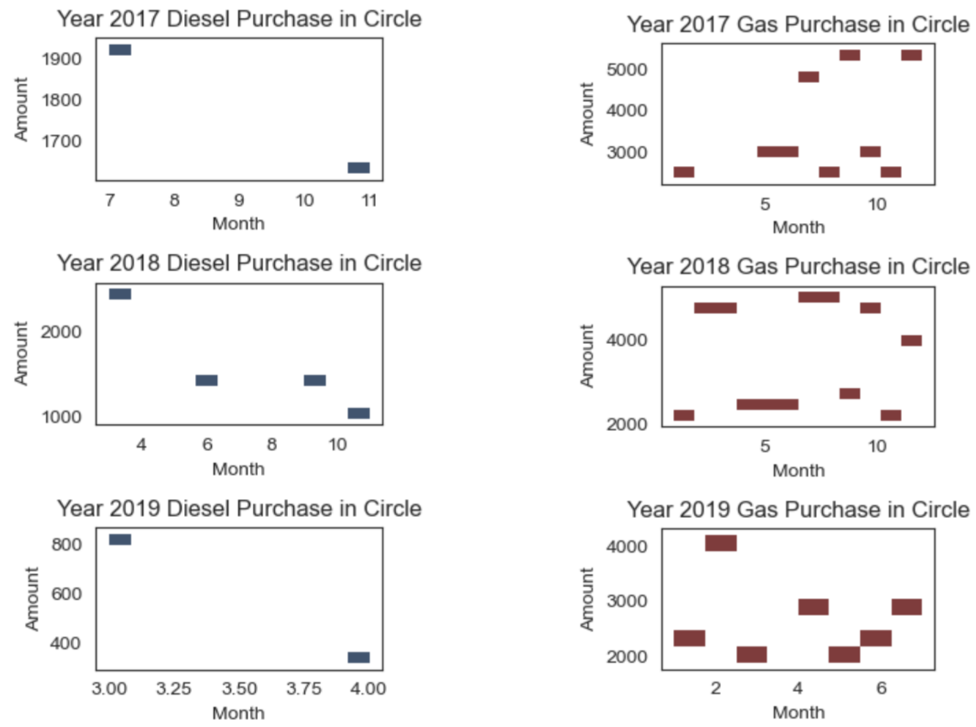
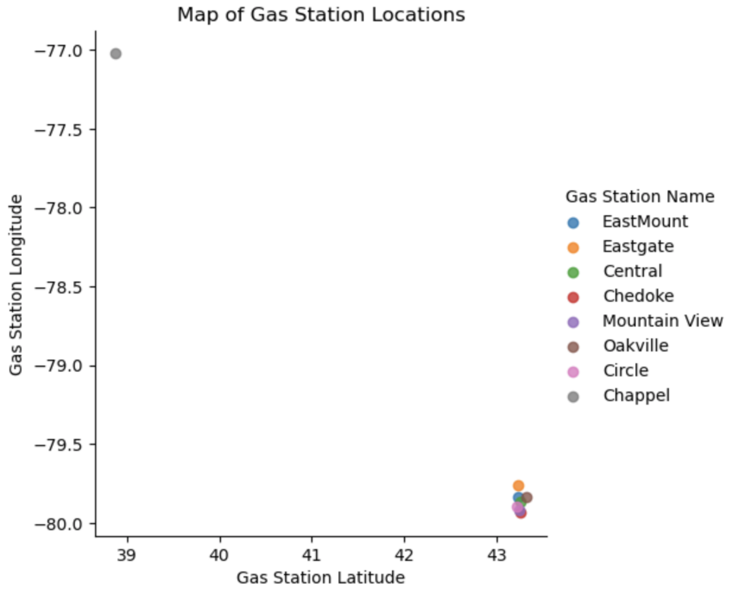


Figure A.6 Diesel and Gas Purchase in Circle

Figure A.7 Diesel and Gas Purchase in Chappel

Appendix B Geographical distribution of gas stations

Set gas stations' latitude as x-label, and gas stations' longitude as y-label, we see that 7 out of 8 stations are tightly connected to each other, they all fall into a specific range. However, gas station Chappel is distinctively located, far away from the rest of other stations.



If we exclude gas station Chappel and only explore the rest 7 stations, we see that there is no specific pattern for the geological distribution. Relatively speaking, Eastgate and Oakville are far away from the rest of the 5 stations.